Lecture 5: Statistical Methods Part II

September 10, 2015

Review from Last Time (I)

- Theory of probability tells us how often to expect a given outcome if we know the underlying distribution that defines the physical process
  - Probability Density Function (pdf) $f(x;\theta)$: prob that x lies between x and $x + dx$.

$$\int_{-\infty}^{\infty} f(x,\theta)dx = 1$$

- Statistical methods tell us how to estimate true quantities given a set of measurements
  - Needs a *model* to define the pdf
  - Statistical questions particle physicists ask:
    - What is our best estimate of a given parameter of our model (or of a quantity that depends on the parameters of our model)?
    - How precise is our estimate of that parameter?
    - Is the data consistent with our model?
    - What range of parameters θ_i are allowed?
    - Can we distinguish between two models and choose which is best?

Reminder: Log Likelihood

- To estimate parameter(s) θ, maximize the likelihood
- Usual technique to find maximum, set derivative equal to zero
- Easier to maximize than $\ln \mathcal{L}$

$$
\begin{aligned}
\frac{\partial \ln \mathcal{L}}{\partial \theta} &= \frac{\partial}{\partial \theta} \ln \prod_{i=1}^{n} \mathcal{L}_i \\
&= \frac{\partial}{\partial \theta} \sum_{i=1}^{n} \ln \mathcal{L}_i \\
&= 0
\end{aligned}
$$

- If several θ_i can minimize with respect to each

Reminder: Connecting the Log Likelihood to the χ^2

$$\mathcal{L}(x;\theta) = \prod_{i=1}^{n} \mathcal{L}_i$$

- Reminder from previous page: For Gaussian case

$$\ln \mathcal{L} = -\sum_i \frac{(x_i - \mu)^2}{2\sigma^2} + const$$

- Compare this to

$$\chi^2 \equiv \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{\sigma^2}$$

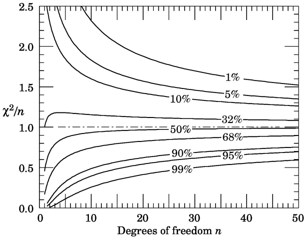
- By inspection

$$\chi^2 = -2 \ln \mathcal{L}$$

- In special (but important) case of Gaussian distributions, the χ^2 and $\ln \mathcal{L}$ are related

The Uncertainty on the estimate of θ

- χ^2 calculates distance squared (in units of σ between measured distribution and prediction of the model
- "Expect" $\chi^2/N \sim 1$ if model is good
- Probability that χ^2/N is larger than a specific value as a function of n:



- From Gaussian case, can relate $-2\ln\mathcal{L}$ to χ^2
- Uncertainty on parameter θ estimated by by find values where $-2\ln\mathcal{L}$ increase by 1 unit

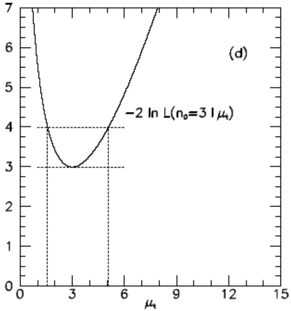


Figure from R. Cousins,
Am. J. Phys. 63 398 (1995)

Goodness of Fit

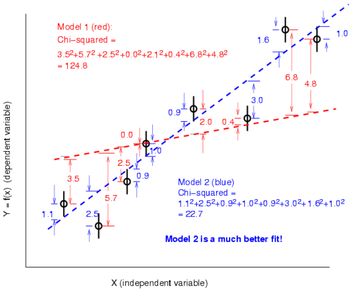
- But estimates of parameters and the uncertainties only makes sense if the model (pdf) used to determine the estimate is correct
- It's not trivial to determine whether a model is good
- Generically, we call parameter determination "fitting" the data
- Determination of whether a model is correct means asking whether the data is consistent with coming from the proposed pdf
  - This is called determining the "goodness of fit"

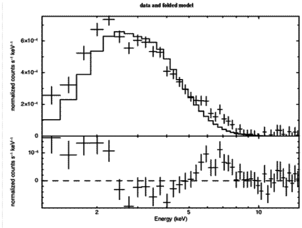
χ^2 Test of Goodness of fit

- Measures distance (in uncertainty space) between data and model

$$\chi^2 = \sum_i \frac{(y(x_i) - f_{model}(x_i))^2}{{\sigma_{y_i}}^2}$$

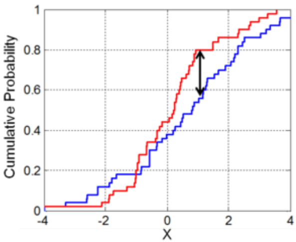
- Only works when uncertainties are symmetric (Gaussian limit)
- Insensitive to whether data is above or below the prediction
  - Not ideal for flagging systematic deviations in shape
  - Eg, for lower plot to the right, χ^2 test gives less discrepancy than you would notice by eye
- Use plot from page 5 to determine probability that data would have a χ^2 value at or larger than what we have measured





Shape Dependent Tests: Kolmogorov-Smirnov (KS)

- Test designed to determine consistency between two datasets
- Can either be two separate measurements (eg do women and men get the same number of colds per year) or a measurement and a prediction (eg real data and MC data)
  - Two samples not required to have same number of events
- No assumptions about the shape of the pdfs, only require that the statistical size of the samples be "large enough"
- For each sample, order data in measured variable (x) and calculate cumulative probability distribution (cpd)



- Measure difference between the two cpd's as function of x
- Identify the largest difference D
- Probability that the two distributions come from the same fundamental pdf:

$$P(D) = 2\sum_{j} = 1^{\infty}(-1)^{j-1}e^{-2j/D^2}$$

see http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm for a good description

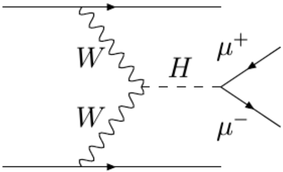
Constructing a model

- To apply any statistical test, we need a model for the data
- These models often complicated to construct.
- Require knowledge of
  - Underlying physics distribution
  - How detector affects the measurement
  - Backgrounds present in the sample
- Kyle Cranmer in 2014 CERN summer school lectures calls this "The Scientific Narrative"
  - Describes 3 common narratives
    - Monte Carlo Simulation narrative
    - Effective modeling narrative
    - Data driven narrative
  - His slides are great. I've included them here.

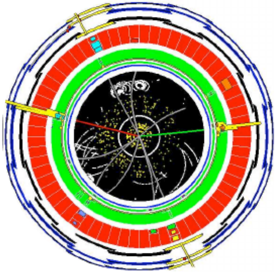
The Monte Carlo Simulation narrative

Let's start with "the Monte Carlo simulation narrative", which is probably the most familiar

$P =$



$\oplus$



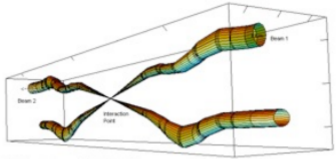
1) The language of the Standard Model is Quantum Field Theory

Phase space Ω defines initial measure, sampled via Monte Carlo

$$P = \frac{|\langle f|i\rangle|^2}{\langle f|f\rangle\langle i|i\rangle}$$

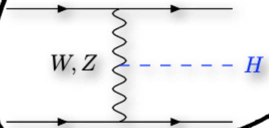
$$P \to L\sigma$$

$$d\sigma \to |\mathcal{M}|^2 d\Omega$$



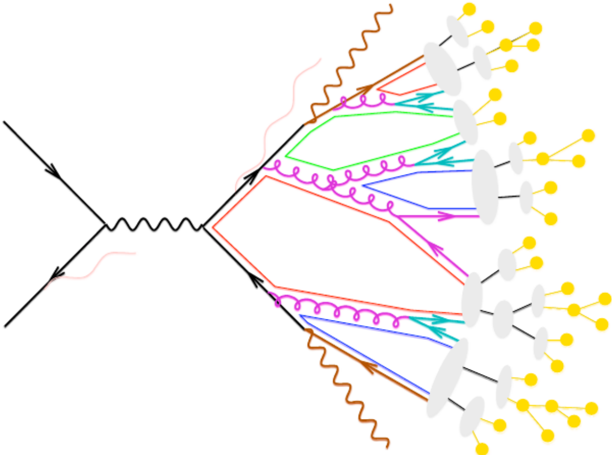
Relative beam sizes around IP1 (Atlas) in collision

$$\begin{aligned}
\mathcal{L}_{SM} = \; & \underbrace{\frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G_a^{\mu\nu}}_{\text{kinetic energies and self-interactions of the gauge bosons}} \\
& + \underbrace{\bar{L}\gamma^\mu(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)L + \bar{R}\gamma^\mu(i\partial_\mu - \frac{1}{2}g'YB_\mu)R}_{\text{kinetic energies and electroweak interactions of fermions}} \\
& + \underbrace{\frac{1}{2}\left|(i\partial_\mu - \frac{1}{2}g\tau\cdot\mathbf{W}_\mu - \frac{1}{2}g'YB_\mu)\,\phi\right|^2 - V(\phi)}_{W^\pm, Z, \gamma, \text{and Higgs masses and couplings}} \\
& + \underbrace{g''(\bar{q}\gamma^\mu T_a q)\,G^a_\mu}_{\text{interactions between quarks and gluons}} + \underbrace{(G_1\bar{L}\phi R + G_2\bar{R}\phi_c L + h.c.)}_{\text{fermion masses and couplings to Higgs}}
\end{aligned}$$



The simulation narrative

2) a) Perturbation theory used to systematically approximate the theory.
b) splitting functions, Sudokov form factors, and hadronization models
c) all sampled via accept/reject Monte Carlo **P(particles | partons)**

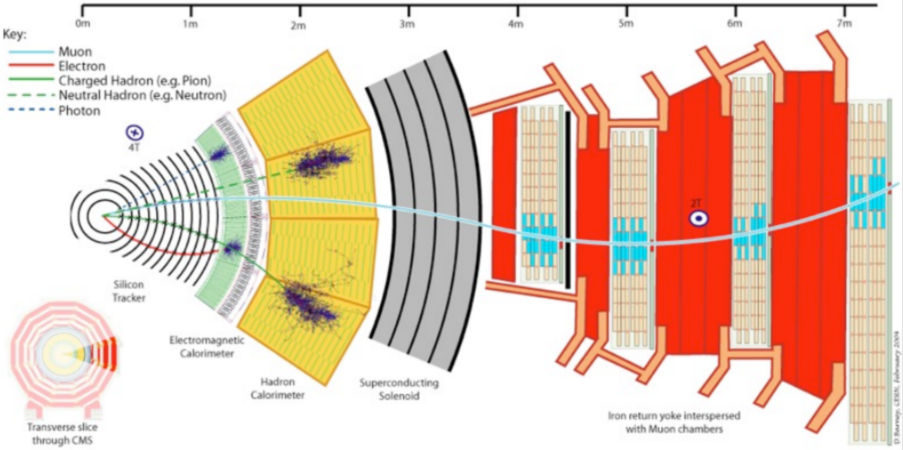


- hard scattering
- (QED) initial/final state radiation
- partonic decays, e.g. $t \to bW$
- parton shower evolution
- nonperturbative gluon splitting
- colour singlets
- colourless clusters
- cluster fission
- cluster $\longrightarrow$ hadrons
- hadronic decays

The simulation narrative

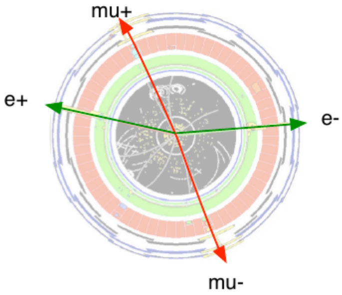
Center for Cosmology and Particle Physics

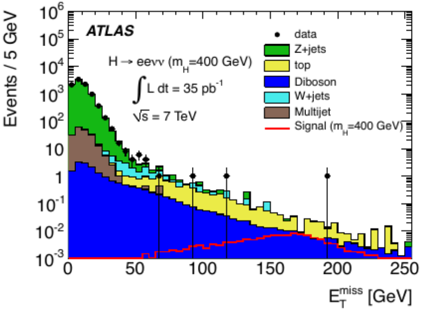
3) Next, the interaction of outgoing particles with the detector is simulated.
Detailed simulations of particle interactions with matter.
Accept/reject style Monte Carlo integration of very complicated function
P(detector readout | initial particles)



The simulation narrative

4) From the simulated response of the detector, we run reconstruction algorithms on the simulated data as if it were from real data. This allows us to look at distribution of any observable that we can measure in data.
P(observable | detector readout)



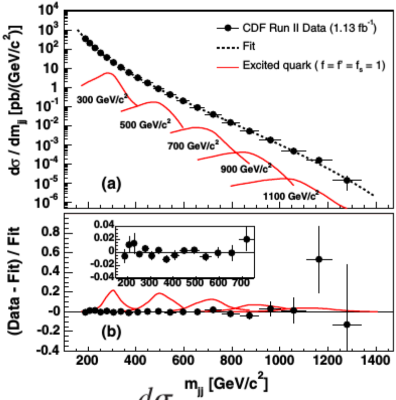


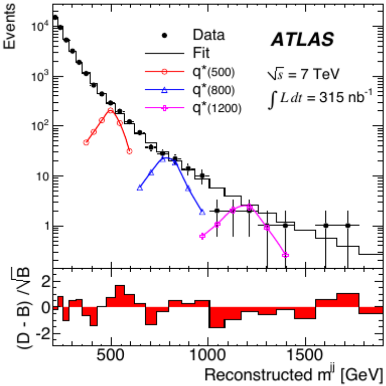
The Effective Model Narrative

In contrast, one can describe a distribution with some parametric function

- "we fit background to a polynomial", exponential, ...
- While this is convenient and the fit may be good, the narrative is weak

PHYSICAL REVIEW D **79,** 112002 (2009)



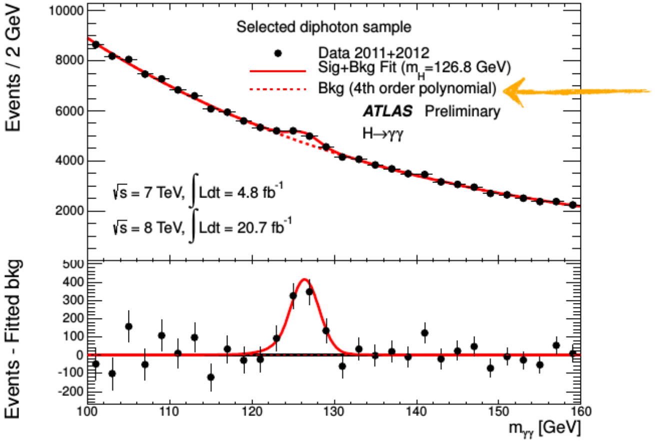


$$\frac{d\sigma}{dm_{jj}} = p_0(1-x)^{p_1}/x^{p_2+p_3\cdot\ln(x)}, \qquad x = m_{jj}/\sqrt{s},$$

The Effective Model Narrative

In contrast, one can describe a distribution with some parametric function

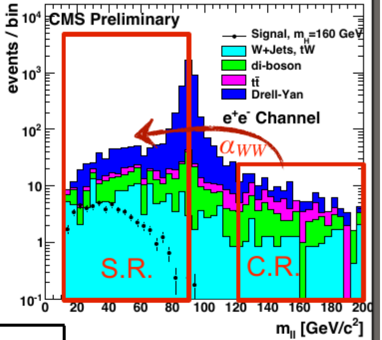
- "we fit background to a polynomial", exponential, ...
- While this is convenient and the fit may be good, the narrative is weak

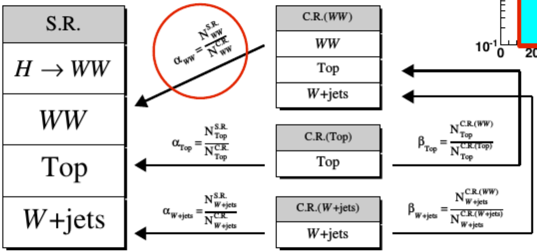


The Data-driven narrative

Regions in the data with negligible signal expected are used as control samples

- simulated events are used to estimate extrapolation coefficients
- extrapolation coefficients may have theoretical and experimental uncertainties





Notation for next slides:

\# in S.R. $\to n_{\mathrm{on}}$

\# in C.R. $\to n_{\mathrm{off}}$

$\alpha_{\mathrm{WW}} \to \tau$

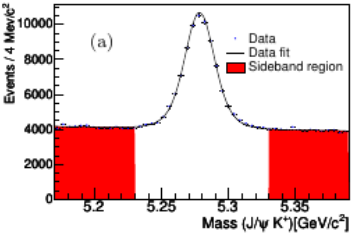
Figure 10: Flow chart describing the four data samples used in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ analysis. S.R and C.R. stand for signal and control regions, respectively.

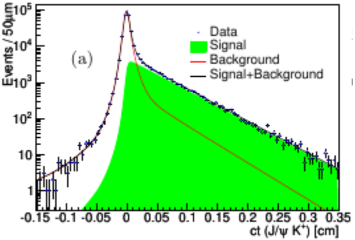
Example of Data Driven Narrative

- CDF measurement of B^+ lifetime

$$B^+ \to \psi K^+, \;\; \psi \to \mu^+ \mu^-$$

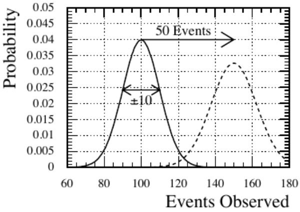
- Invariant mass of $\mu^+\mu^- K$ system shows peak at B^+ mass, plus flat background
- Decay position determined from fit to $\mu^+\mu^- K$ vertex (distance from primary vertex)
- Difficult to model distribution of vertex position of background from MC
- Instead use sidebands of mass distribution
- Unbinned likelihood fit to combined likelihood for signal and background





Hypothesis Testing: The Likelihood Ratio

- Experiments typically have background in addition to signal
- How do we know if there is a significant signal "on top of" the background?
- Standard technique: Construct pdfs under two hypotheses
  - Background only hypothesis
  - Signal plus background hypothesis
- Ask whether the data are sufficient to distinguish between the two



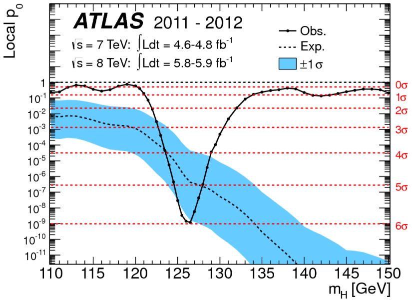
The Neyman-Pearson Lemma

- Suppose we have two hypotheses
  - H_0: Background only (Null Hypothesis)
  - H_1: Signal plus background
- Define α as the probability that we wrongly reject the Null Hypothesis for a given set of measurements
- We want to define the region W for which we minimize the probability of wrongly accepting H_0 if H_1 is true
- The region that minimizes the probability of wrongly accepting H_0 is a contour in the likelihood ratio

$$\lambda(\vec{N}) = \frac{\mathcal{L}(\vec{N}|H_{S+B})}{\mathcal{L}(\vec{N}|H_B)}$$

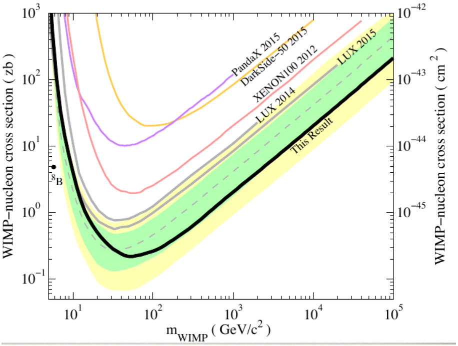
- Can define a "p-value" that relates the measured likelihood ratio to the fraction of the time that H_0 would have a likelihood ratio at least as large as the data.

Example: p-value and Higgs Discovery



- Local p-value vs m_H
- Dotted line is expected p-value for a SM Higgs with that mass
- Warning: "Look-elsewhere" effect
  - When asking how likely something is, must take into account how many places you looked!

Example: LUX Dark Matter Exclusion Limits

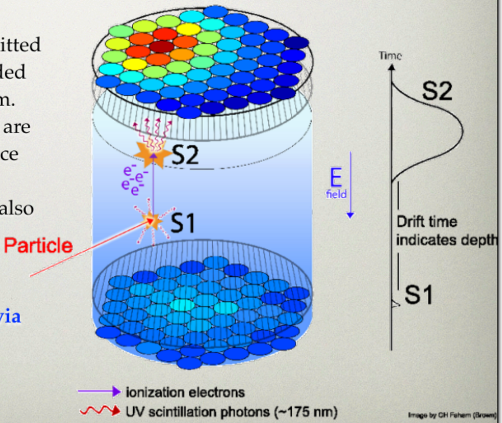


- Express measurement as contour in space of relevant parameters (here, 90% c.l.)
- Show expected limit as well as measured one
  - If measurement far from expected sensitivity, you should worry

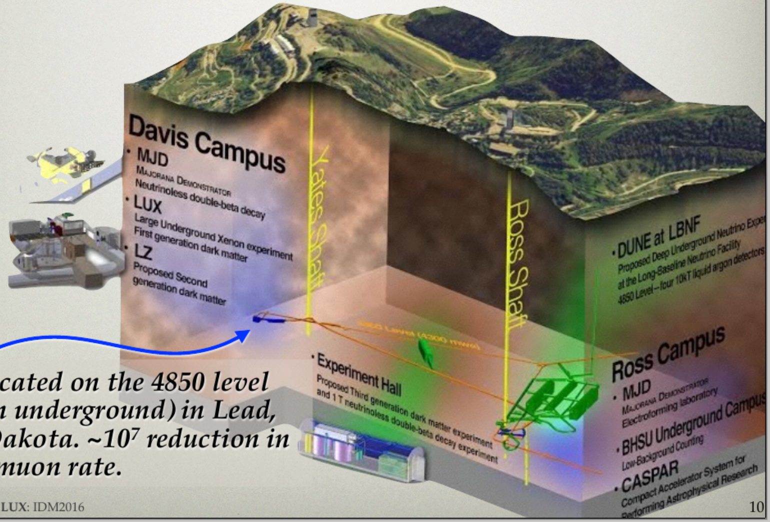
Next 3 slides taken from A. Manalaysays' talk presenting lates results July 2016

Detection technique

- LUX is a dual-phase time projection chamber (like most other liquid-noble DM experiments); essentially a cylinder of LXe.
- Primary scintillation light ("S1") is emitted from the interaction vertex, and recorded by an array of PMTs on top and bottom.
- Electrons emitted from the interaction are drifted by an applied field to the surface and into the gas, where they emit proportional scintillation light ("S2"), also recorded by the PMTs.
- This design permits:
  - 3-D localization of each vertex.
  - Identification of multiple scatters (via S2 count).
  - ER/NR discrimination (via S2/S1)
  - Sensitivity to single ionization electrons.



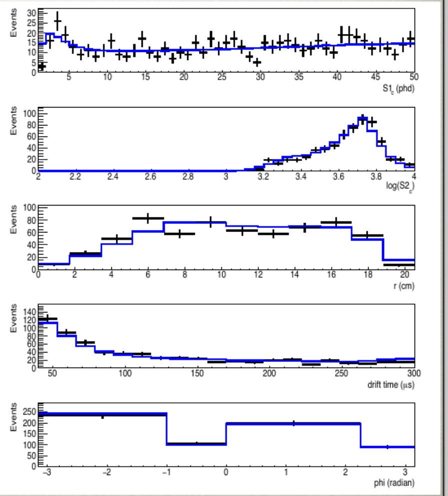
Sanford Underground Research Facility



***LUX**, located on the 4850 level (~1.5 km underground) in Lead, South Dakota. ~10^7 reduction in cosmic muon rate.*

Profile Likelihood Analysis

- Data are compared to models in an un-binned, 2-sided profile-likelihood-ratio (PLR) test.
- 5 un-binned PLR dimensions:
  - Spatial: r, ϕ, drift-time (raw-measured coordinates)
  - Energy: S1 and $\log_{10}$(S2)
- 1 binned PLR dimension:
  - Event date
- The data in the upper-half of the ER band were compared to the model (plot at right) to assess goodness of fit.



Systematic Uncertainties

- Systematic uncertainties an art rather than a science
- Need to estimate how wrong your hypothesis might be due to
  - Mis-modeling of detector response (calibration or resolution)
  - Variations in conditions (eg temp, pressure)
  - Mis-modeling of physics processes (signal or background)
- Such uncertainties aren't necessarily Gaussian (but we often assume that they are)
- Typically quote them separately from statistical uncertainties

A New Trend: Profiling of Systematic Uncertainties'

- Itemize sources of systematic uncertainty (usually lot's of them)
  - Often called "nuisance parameters"
- Attempt to write pdf's for how systematic uncertainties affect the measurement
  - Use simulated data to determine pdf as function of quantity you are measuring
- Include these uncertainties in likelihood function
  - Usually have to assume Gaussian uncertainties
- "Profiling" is fancy word for including these in the fit
- Idea is to use the data to constrain these parameters
- Powerful but also dangerous